

Quantifying the linguistic persistence of high and low performers in an online student forum

Carissa Myers,¹ Elizabeth Fox,² Adrienne Traxler,¹ and A. Gavrin³

¹*Department of Physics, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH, 45345*

²*Department of Psychology, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH, 45345*

³*Department of Physics, Indiana University Purdue University Indianapolis, 402 N. Blackford St., Indianapolis, IN, 46202-3217*

This work uses recurrence quantification analysis (RQA) to analyze the online forum discussion between students in an introductory physics course. Previous network and content analysis found differences in student conversations occurring between semesters of data from an introductory physics course; this led us to probe which concepts occur and persist within conversations. RQA is a dynamical systems technique to map the number and structure of repetitions for a time series. We treat the transcript of forum conversations as a time series to investigate and apply RQA techniques to it. We characterize the forum behaviors of high and low scoring students, such as their percentage of recurring topics and persistence of discussing a topic over time. We quantify how high scoring and low scoring students use online discussion forum and test whether different patterns exist for these groups. This work is the first adaptation of recurrence quantification methods from the field of psychology for physics education research. Using RQA, there was not a general, observable difference in how the two different groups, high- and low-scoring students, used the forum; however, there were differences when focusing in on and comparing one high-scoring student and one low-scoring student. This technique has the potential for analyzing other PER data such as interviews or student discussions.

I. INTRODUCTION

Recurrence quantification analysis (RQA) was performed on conversations from an online discussion forum of an introductory physics course. Recurrence quantification analysis is a technique that characterizes recurrent structures in a time series by quantifying the number and duration of repetitions [1]. The goal of this work is to use RQA to analyze the posts of high and low grade students within the online forum.

The motive for this study stems from prior work using social network analysis to examine three semesters of introductory physics online forum data [2]. In that work, network position correlated with course grade in the first and third semesters of data, but not the second semester. Network analysis could not be used to explain this variation. The only difference between semesters was the use of anchor threads, small activities given to students to encourage engagement in conversation within the online discussion forum. This led us to believe anchor threads might be one source of the difference between semesters because they were not used in the second semester. Content analysis was used to investigate the use of anchor threads in the first semester of data, and demonstrated differences between the use of anchor and non-anchor threads (student conversations that were not related to anchor tasks) [3]. The content analysis tool detected shifts in sentence types, but had limitations in terms of detecting themes in student discussions and reasons behind the behavior such as whether anchor tasks encouraged productive conversations. Since the content analysis tool observed surface-level behaviors and patterns of student discourse, recurrence quantification analysis was introduced to look further into the patterns and behaviors of the text.

Because recurrence quantification analysis is a method for analyzing complex dynamical systems and student discourse can be considered a complex dynamical system, this leads us to look for recurrent structures within student conversations. In psychology, RQA is used to study many different types of communication, from medical consultation to cockpit discourse to professional interviews, to understand more about the patterns within these conversations [4, 5]. Angus and collaborators have used recurrence quantification analysis to analyze a conversation between a talk show host and guests as well as doctor-patient consultations [6, 7]. RQA showed complex dynamics such as one person trying to change the subject and another person dragging it back to their preferred topic. Though it was designed for other types of dynamical systems, this work demonstrated that recurrence quantification analysis can be used to learn more about the patterns in conversation.

The goal of our research was to use recurrence quantification analysis to identify patterns in conversations from an online discussion forum of introductory physics students. The long-term goal would be to connect patterns in conversations to the prior network analysis, and identify patterns of conversations associated with groups of students. For instance, how do the conversation patterns compare between a student

who has high centrality and high grade, versus a student who has a lower grade and lower centrality in the network? This work takes the first step by analyzing the forum discussions for groups of high and low grade students. For PER, this work is important because it is introducing a new technique to analyze conversations, which could be extended to other types of text. This paper explores how RQA can be used to show patterns of student conversations across a semester, and the results suggest that RQA can provide useful insight into the structure of student conversations.

II. METHODS

A. Course Context and Data

Recurrence quantification analysis was performed on an online discussion forum of an introductory, calculus-based physics course, which was a lecture with active learning [2, 8, for details]. There were 173 total students within the course, and 156 students as well as the instructor posted in the forum. There were 936 threads and 2376 reply comments. Students could earn up to 5 percent extra credit on their final grade by participating in the forum.

The class consisted of 21% women and 79% men. Detailed racial and ethnic demographics are unavailable for the class. The institution is a large, urban, public university whose students were 72% White, 10% African American, 6% Hispanic or Latino, and other groups (including international students) 4% or less [2].

These data were forum posts from one semester of a course. Each post included a student identifier and timestamp. The length of posts varied, ranging from a single word (e.g., “Thanks”) to multiple sentences. RQA was performed for 6 low grade students and 6 high grade students, or 12 students total. These students were chosen based on being in the top and bottom 5 percent of the class as well as posting more than 10 posts within the forum. The last restriction is to ensure that enough words and utterances are present to be analyzed by the RQA code.

B. Recurrence Quantification Analysis

The text from the student forum was processed so that any two posts, also termed “utterances,” could be compared for similarity. We follow the procedure of Angus *et al.* [6]; see that work for details. This processing starts with selecting a student and pulling their posts from the transcript. The next step removes stop words, which are commonly used words that hold no meaning such as “and,” “in,” and “a.” Posts were then broken into sentences using punctuation, yielding a character vector consisting of one sentence per entry. The sentence vectors were broken into windows of three sentences. Punctuation and capitalization were removed as the next part of the processing step. A list of unique terms T was built

for the student, ordered by the posting date (beginning of the semester to the end of the semester). An occurrence vector was created, which is the frequency of all unique words across all the windows. Then a co-occurrence matrix was built, where each element represents the frequency of any pair of terms co-occurring in the same window. These values were used to calculate the similarity of terms.

Similarity can be calculated using the number of sentence windows, the occurrence vector, and co-occurrence count. The semantic similarity index of terms i and j is

$$S(t_i, t_j) = \frac{P(t_i, t_j) \times P(\bar{t}_i, \bar{t}_j)}{P(\bar{t}_i, t_j) \times P(t_i, \bar{t}_j)} \quad (1)$$

$$P(t_i, t_j) = c_{ij} / N \quad (2)$$

$$P(\bar{t}_i, \bar{t}_j) = \begin{cases} 1, & \text{if } O_i + O_j = c_{ij} + N \\ \frac{N - O_i - O_j + c_{ij}}{N}, & \text{otherwise.} \end{cases} \quad (3)$$

$$P(t_i, \bar{t}_j) = \begin{cases} 1, & \text{if } O_i = c_{ij} \\ (O_i - c_{ij}) / N, & \text{otherwise.} \end{cases} \quad (4)$$

$$P(\bar{t}_i, t_j) = \begin{cases} 1, & \text{if } O_j = c_{ij} \\ (O_j - c_{ij}) / N, & \text{otherwise.} \end{cases} \quad (5)$$

N is the number of sentence windows, O_i is the occurrence count of term i , and c_{ij} is the co-occurrence count of terms i and j [9]. Conceptually, $P(t_i, t_j)$ can be thought of as the probability that both terms i and j appear in a window. $P(\bar{t}_i, \bar{t}_j)$ is the probability that neither term i nor j appears. $P(t_i, \bar{t}_j)$ is the probability that term i appears without term j , and $P(\bar{t}_i, t_j)$ is the equivalent calculation for finding j without i . The similarity of terms increases if they tend to occur in the same windows (and are both absent in other windows). Similarity decreases when two terms often appear separately from each other.

Once the similarity was calculated using windows, the next step was to define an utterance. In this work, an utterance was a single post, which could consist of one word to one sentence to multiple sentences. This reorganization of sentences into utterances is a common process step [9]. To analyze the concepts within the utterances, a key term list, K , is built. Different criteria, usually frequency-based, can be used to create the list of key terms. In the calculations reported here, we used the full unique words list (so $K = T$). The similarity matrix calculated consists of the similarity between each key term (K) and each term from the term list (T). Then a Boolean matrix B is built where each element indicates whether each term from the full list T is present in an utterance. Finally, a feature matrix V is built by matrix multiplication of the similarity matrix S and Boolean matrix: $V = SB$. The dot product of any two columns within V gives the similarity of those two utterances.

A recurrence plot can be built now that the similarity between utterances has been calculated. A recurrence plot is a two-dimensional plot, where both of the axes represent a time series [10]. The recurrence plot visualizes the similarity between utterances at different periods of time [6]. On each

plot, there is a similarity scale ranging from 0 to 1, which shows a continuum of shading representing how similar the terms are between two utterances. This shading is also due to the fact that each utterance may have a different number of words from the key terms list, K . A high similarity means there are more words in the window that match the subset of key terms (higher frequency of recurring words) whereas a low similarity means there are fewer words in the window that match the subset of key terms (a more diverse choice of words).

To gain further insight into the behaviors and patterns associated with recurrence plots, four measures—recurrence rate, determinism, longest diagonal line, and average diagonal line—were chosen. Equations below are adapted from [11], using additional code from the `crqa` package in R [12, 13]. To calculate these statistics, two parameters, similarity threshold and diagonal line threshold (explained further below), were set. A similarity threshold means for two utterances to be considered similar, they must have a similarity value equal to or above a set value. In this work, the similarity threshold was set to 0.5.

Recurrence rate, RR , represents the percent of recurring points falling above the similarity threshold [1]. This is calculated using Eq. (6), using the inverse of the number of utterances squared multiplied by the sum of all the points above the similarity threshold:

$$RR = \frac{2}{N^2} \sum_{i,j=1}^N R_{i,j}, \quad (6)$$

where $R_{i,j}$ is one for all points above the threshold. A percent recurrence rate results when multiplying RR by 100. A high recurrence rate means that the students were repeating many words or phrases, whereas a low recurrence rate means the students were using a larger variety of words with fewer repetitions.

Determinism is the fraction of recurring points forming diagonal lines on the recurrence plot [1]. To define a diagonal line, there must be some minimum number of points in a row; this parameter is the diagonal line threshold. In this work, it was set to 2. Equation (7) was used to calculate determinism, which is the number of points in diagonal lines divided by the number of points above the similarity threshold.

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (7)$$

Here, l_{min} is the diagonal line threshold, and $P(l)$ is the histogram of the lengths of the diagonal lines. A diagonal line indicates that over a period of time, the student discussed a sequence of concepts or repeated a phrase that had already been mentioned before. Therefore, determinism quantifies how persistent a system is, where the system here is a student's discussion in the online forum.

To understand more about the diagonal structures, the longest diagonal line length (L_{max}) is calculated over all diagonals (lengths l_i). The main diagonal, where all points are

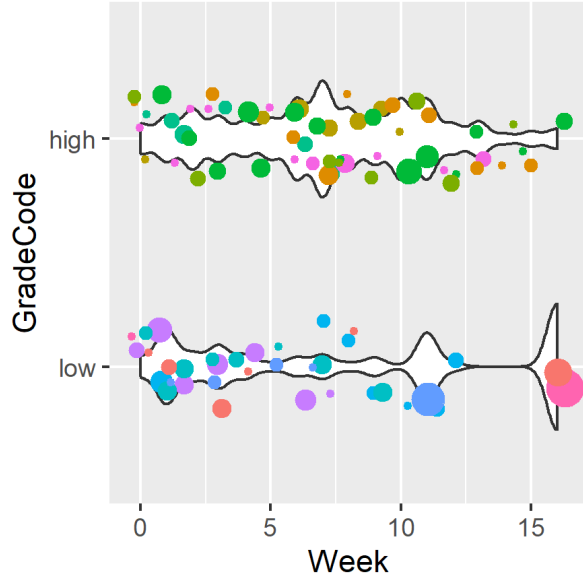


FIG. 1. Posting density as a function of week in the semester, grouped by high- and low-scoring students. Wider places in the density plot show a higher frequency of posting at those times. Individual posting data is superimposed (one student per color), with point size scaled by frequency and a small random jitter to avoid overlap.

1 by definition, is omitted from this calculation:

$$L_{max} = \max(l_i; i = 1, \dots, N_l) \quad (8)$$

The average diagonal length is the average of the lengths of the diagonal lines, calculated using Eq. (9). This average is an indication of how often a number of words or a phrase are sequentially repeated.

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (9)$$

III. RESULTS

Students in the high-scoring group posted 28 ± 19 (mean \pm one standard deviation) times during the semester, with their sentence count averaging 71 ± 54 . Students in the low-scoring group averaged 30 ± 6 posts and 73 ± 62 sentences. On volume alone, there is much more variation within groups than between them. Figure 1 summarizes when this posting occurred. There was some tendency for low-scoring students to cluster their conversation at the beginning and end of the semester, while the group of high-scoring students collectively had a more steady output.

The recurrence plots for one high grade and one low grade student are shown in Fig. 2. These students were chosen based on their number of utterances and recurrence rates,

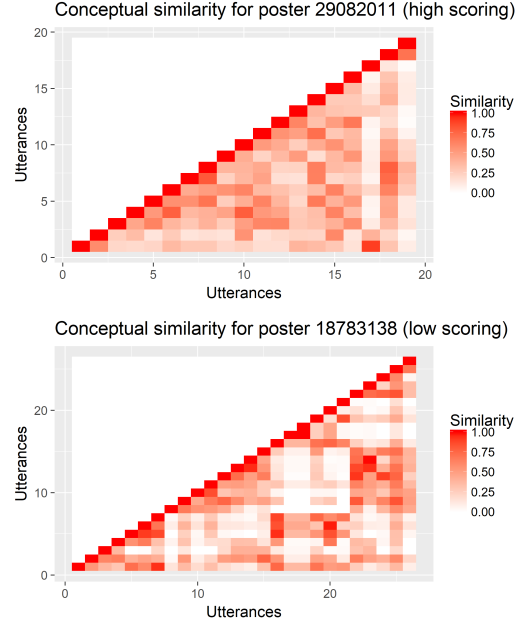


FIG. 2. Recurrence plots for a high-scoring student (top) and low-scoring student (bottom). The x - and y -axis both show utterances (posts) during the semester. Each square is shaded by the amount of similarity between the two utterances. (For example, $x = 10$ and $y = 2$ shows the similarity between the tenth and second posts of the student.) The similarity scale on the plots ranges from zero to one, with zero meaning there is no similarity and one meaning maximum similarity (complete overlap of terms) across both utterances.

which were close to the midpoint for each of their representative groups. In this example pair, there was a different number of utterances for the high- and low-scoring students (19 vs. 26 posts). The recurrence rates for the two students were similar (22.2% vs. 23.1%), meaning that they tended to return to talking about previous ideas at the same rate (though not always at the same time). The percent determinism was larger for the high-scoring student (57.5 vs. 44.9). This suggests the high-scoring student, when picking up a prior line of conversation, may have stayed with it for longer. There were corresponding slight differences in the diagonal lengths with the high-scoring student having slightly longer lines (2.3 vs. 2.2 average, and L_{max} of 4 vs. 3).

To characterize the recurrence statistics for the high/low scoring subset of students, Table I compares averages for recurrence rate, determinism, average diagonal length, and maximum diagonal. There was no general pattern in regard to the density of recurrence plots for high- and low-scoring students.

In exploratory analyses, we varied the similarity threshold across 10 equally-spaced intervals from 0.2 (less stringent) to 0.65 (more stringent). A statistical test of variance in each measure of RQA indicated that RR, DET, average line length, and maximum line length varied as a function of similarity threshold. Specifically, RQA measures increased

TABLE I. Group-level descriptive statistics for each RQA metric, separated by high- and low-scoring students. The similarity threshold was 0.5, and the diagonal line threshold was 2.

RQA Metrics	High-Scoring		Low-Scoring	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recurrence rate (RR %)	27	12	27	7
Determinism (DET %)	48	16	43	9
Average diagonal line length	2.5	0.2	2.4	0.2
Maximum diagonal line length	5	2	4.5	1.4

Note: M: Mean; SD: Standard deviation

for less stringent similarity criteria. However, this pattern was consistent across high and low scoring students. In sum, the results of each 2 (grade code) \times 10 (similarity threshold) way ANOVA indicated a significant main effect of similarity threshold, a non-significant main effect of grade code, and a non-significant interaction between the two across each RQA measure.

IV. DISCUSSION

The main objective of this study was to demonstrate how RQA can be applied to PER and give one kind of model of how it can be used in PER. To achieve this, we focused on the analysis of the recurrence plots and RQA measures for high- and low-scoring students. We gained insight into the posting behaviors of these individuals when comparing one another, but there were not general, observable differences in posting behaviors between the larger groups of high- and low-scoring students using recurrence quantification analysis. When choosing one high-scoring student and one low-scoring student, RQA was able to show the high-scoring student infrequently posted, but was typically more persistent in conversation across the semester. In terms of recurrence rate, the high-scoring student was a little more focused than the low-scoring student because in order for them to have a similar recurrence rate, the low-scoring student posted more in the forum than the high-scoring student. Therefore, RQA allows for the texture and conceptual content of different posting behavior between students to be analyzed.

Educational technology or online learning researchers have analyzed student forum conversations using a variety of frameworks as well [14]. That work primarily uses quantitative content analysis, which we had used previously [3] and found insufficient to describe patterns of interest in this data. We expect RQA might complement qualitative methods. For example, Bruun and collaborators [15] used qualitative discourse analysis together with network analysis to extract themes from a class discussion. RQA could add information about the time structure of the conversation and help to identify which conversational bids were taken up by classmates. Our code for calculating conceptual similarity and re-

currence quantification statistics is available online [16].

To cut to a small “proof of concept” subset of the class, in this work we focused on high- and low-scoring students in the class. However, to generate readable recurrence plots, we also required that students posted at least ten times during the semester. This threshold may have selected students who were atypical in some fashion. For example, low-scoring students tended to have low centrality in the forum network [2]. While this is not the same as a low posting rate, it is possible that the ten-post requirement restricted our analysis to an unusually talkative set of low-scoring students. To look for these effects, RQA statistics could be generated for all students in the class, at which point an ANOVA or other methods could be used to look for grade-based variation in RQA values.

There were also some underlying behaviors and patterns not identified in this study that could be further learned about using RQA. RQA is a very powerful tool, which allows many parameters to be set. For example, one result from this work is the similarity threshold has a significant impact on the results of the RQA measures. Similarity threshold is not the only specification that could be changed that would impact the results. The key terms subset, K , could be changed from most frequent terms to a list of physics terms or any list of interest (for example, words from the anchor tasks). Using a built list of physics terms for K might find that high-scoring students are more focused and persistent in discussing physics than low-scoring students. By building new subsets, more information could be gained about what topics students are focusing on, and how they are using specific topics within the forum. Future work would include creating new subsets, and analyzing if and how the posting behaviors change for students.

Other future work would include pulling out what concepts recur over time and when those concepts recur over time. Examples include: (1) Are high-scoring students discussing the topics when they come up in the class or are they discussing the topics before they come up in class? (2) Are low-scoring student discussing the topics primarily right before the exams? (3) Do high-scoring students discuss the connections between topics as the semester progresses? Learning what topics recur over time, and when those topics recur, is useful information because arguments could be made regarding why a student has a higher or low grade or is more or less central to the network. More importantly, online forum behavior could be identified that may help students be more successful in the course.

ACKNOWLEDGMENTS

Carissa Myers was supported by a PERLOC travel grant in attending this conference.

-
- [1] J. Webber, Charles L. and J. P. Zbilut, Recurrence quantification analysis of nonlinear dynamical systems, in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, edited by M. A. Riley and G. C. Van Orden (2005) pp. 26–94.
- [2] A. Traxler, A. Gavrin, and R. Lindell, Networks identify productive forum discussions, *Physical Review Physics Education Research* **14**, 020107 (2018).
- [3] C. Myers, A. Traxler, and A. Gavrin, Content analysis of instructor tools for building a learning community, in *2018 Physics Education Research Conference*, edited by A. Traxler, Y. Cao, and S. Wolf (Washington, DC, 2018).
- [4] G. Leonardi, The study of language and conversation with recurrence analysis methods, *Psychology of Language and Communication* **16**, 165 (2012).
- [5] R. Fusaroli, I. Konvalinka, and S. Wallot, Analyzing social interactions: The promises and challenges of using Cross Recurrence Quantification Analysis, in *Translational Recurrences*, Springer Proceedings in Mathematics & Statistics, edited by N. Marwan, M. Riley, A. Giuliani, and J. Webber, Charles L. (Springer International Publishing, 2014) pp. 137–155.
- [6] D. Angus, A. E. Smith, and J. Wiles, Human communication as coupled time series: Quantifying multi-participant recurrence, *IEEE Transactions on Audio, Speech, and Language Processing* **20**, 1795 (2012).
- [7] D. Angus, B. Watson, A. Smith, C. Gallois, and J. Wiles, Visualising conversation structure across time: Insights into effective doctor-patient consultations, *PLOS ONE* **7**, e38014 (2012).
- [8] A. Gavrin and R. S. Lindell, Connecting students’ homework to their participation in a course-based social network, in *2017 ASEE Annual Conference & Exposition* (Columbus, OH, 2017).
- [9] D. Angus, A. Smith, and J. Wiles, Conceptual recurrence plots: Revealing patterns in human discourse, *IEEE Transactions on Visualization and Computer Graphics* **18**, 988 (2012).
- [10] J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, Recurrence plots of dynamical systems, *Europhysics Letters* **5**, 973 (1987).
- [11] Quantification of Recurrence Plots (Recurrence Quantification Analysis), <http://www.recurrence-plot.tk/rqa.php>.
- [12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2017).
- [13] M. I. Coco, R. D. with contributions of James D. Dixon, and J. Nash, *crqa: Cross-Recurrence Quantification Analysis for Categorical and Continuous Time-Series* (2018), R package version 1.0.9.
- [14] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer, Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review, *Computers & Education* **46**, 6 (2006).
- [15] J. Bruun, M. Lindahl, and C. Linder, Network analysis and qualitative discourse analysis of a classroom group discussion, *International Journal of Research & Method in Education* **42**, 317 (2019), <https://doi.org/10.1080/1743727X.2018.1496414>.
- [16] RQA github repository, <https://github.com/atraxler/RQA>.